On the Security of Doubly Efficient PIR

Elette Boyle* Justin Holmgren† Fermi Ma‡ Mor Weiss§

Abstract

Doubly Efficient Private Information Retrieval (DEPIR) enables queries to an externally held database while hiding the identity of the queried indices, strengthening standard Private Information Retrieval (Chor, Goldreich, Kushilevitz, Sudan FOCS'95) with an efficiency requirement that the computational demands of both client and server are sublinear in the database size. The first DEPIR candidate constructions were recently put forth, based on a new type of assumption relating to indistinguishability of moderate-degree polynomials from random functions when given permuted versions of their evaluation graphs (Boyle, Ishai, Pass, Wootters TCC'17 and Canetti, Holmgren, Richelson TCC'17). To aid in the cryptanalytic study of this new assumption, the work of (BIPW TCC'17) put forth a simpler "toy conjecture" variant.

In this note, we present an attack that provably breaks the BIPW TCC'17 toy conjecture. The attack identifies a natural embedding of permuted samples into a higher-dimensional linear space for which permuted polynomial samples will be rank deficient. We note, however, that our attack does not apply to the real assumption underlying the constructions, and thus the candidates still stand. We discuss extensions of the attack and present an alternative "new toy conjecture" for future study.

Similar results were independently obtained by (Blackwell and Wootters, ArXiv'21).

1 Introduction

Private Information Retrieval (PIR) [CGKS95, KO00] schemes are protocols that enable a client to access entries of a database stored on a remote server (or multiple servers), while hiding from the server(s) which items are retrieved.

It is possible to privately retrieve bits of an N-bit database under a variety of cryptographic assumptions, with as little as polylog(N) bits of communication, and with client running time of just $\mathsf{polylog}(N)$. However, Beimel et al. [BIM00] observed that the security guarantee inherently requires the *server's* computation to be $\Omega(N)$ per query. They proposed circumventing this lower bound with a pre-processing stage, and constructed a multi-server preprocessing PIR scheme that is *doubly efficient*. That is, both the client *and* the server perform $o(N)$ per-query computation after an initial preprocessing computation of size $\Omega(N)$.

Achieving doubly-efficient PIR in the *single server* setting remained completely open until the recent independent works of [BIPW17, CHR17] proposed the first candidate constructions. The security of their candidates is based on a new conjecture that *permuted* local-decoding queries (for a Reed-Muller code with suitable parameters) are computationally indistinguishable from uniformly random sets of points.

In service of future cryptanalysis, [BIPW17] proposed a particular simple "toy conjecture" that was inspired by (but formally unrelated to) their SK-DEPIR scheme.

Our Contributions. In this work, we:

- identify an efficient attack that breaks the aforementioned toy conjecture.

[*] IDC Herzliya. Email: eboyle@alum.mit.edu. Supported by ISF grant 1861/16, AFOSR Award FA9550-21-1-0046, ERC Project HSS (852952), and a Google Research Scholar Award.

[†] NTT Research. Email: justin.holmgren@ntt-research.com.

[‡] Simons Institute & UC Berkeley. Email: fermima@alum.mit.edu.

[§] Bar-Ilan University. Email: mor.weiss@biu.ac.il. Supported by the BIU Center for Research in Applied Cryptography and Cyber Security in conjunction with the Israel National Cyber Bureau in the Prime Minister's Office

- put forth a modified toy conjecture, as a simple target for cryptanalysis.
- put forth an extended conjecture that formally implies the existence of an SK-DEPIR with non-trivial efficiency. We believe this extended conjecture is simpler and more amenable to analysis than the assumptions underlying previous SK-DEPIR schemes.

Concurrent Work. Independent and concurrent to this work, Blackwell and Wootters [BW21] observed essentially the same attack on the [BIPW17] toy conjecture, and proposed essentially the same modification to the toy conjecture. They did not extend the toy conjecture to one that implies a SK-DEPIR scheme.

2 Attack on the [BIPW17] Toy Conjecture

2.1 The Conjecture

Towards understanding the security of their DEPIR construction, [BIPW17] proposed the following conjecture.

Conjecture 1. *Let $\mathbb{F}_q$ be a finite field where $q \approx \lambda^2$. Let $p_1, \ldots, p_m$ be random degree-λ polynomials over $\mathbb{F}_q$, for $m = \lambda^{100}$. Let $r_1, \ldots, r_m$ be random functions from $\mathbb{F}_q$ to $\mathbb{F}_q$. Then the following two distributions are computationally indistinguishable, over the choice of random permutation $\pi \leftarrow S_{\mathbb{F}_q \times \mathbb{F}_q}$. Here, elements of each set S_i or T_i appear in canonical sorted order (not ordered by $x \in \mathbb{F}_q$).*

1. *Distribution D_0: Permuted low-degree polynomials: $(P_1, \ldots, P_m)$, for $P_i = \{\pi(x, p_i(x)) : x \in \mathbb{F}_q\}$.*
2. *Distribution D_1: Permuted random functions: $(R_1, \ldots, R_m)$, for $R_i = \{\pi(x, r_i(x)) : x \in \mathbb{F}_q\}$.*

We first restate an equivalent form of the conjecture in a way that is more amenable to explaining our attack.

Definition 2.1 (Function Graphs)**.** *Let $f : X \to Y$ be a a function. The* graph of f, *denoted $G(f)$, is defined as the set*

$$G(f) \overset{\text{def}}{=} \{(x, y) \in X \times Y : y = f(x)\}.$$

Definition 2.2 (Indicator Strings)**.** *Let S be a subset of a universe U. The* indicator of S (in U), *denoted $\mathbf{1}_S$, is defined as the string in $\{0,1\}^U$ with*

$$(\mathbf{1}_S)_i \overset{\text{def}}{=} \begin{cases} 1 & \textit{if } i \in S \\ 0 & \textit{otherwise.} \end{cases}$$

Definition 2.3 (Permutation Action on Strings)**.** *For any sets X and Σ, we let S_X act on Σ^X in the following (standard) way. For any permutation $\pi \in S_X$ and $x \in \Sigma^X$, we have*

$$(\pi \cdot x)_i \overset{\text{def}}{=} x_{\pi^{-1}(i)}.$$

Restated Conjecture 1. *For $b \in \{0,1\}$, define the distribution ensemble D_b as follows:*

1. *Let $\mathbb{F}_q$ be a finite field where $q \approx \lambda^2$.*
2. *For $m = \lambda^{100}$, sample i.i.d. functions $f_1, \ldots, f_m : \mathbb{F}_q \to \mathbb{F}_q$, where:*
   - *If $b = 0$ each f_i is a uniformly random degree-λ polynomial.*
   - *If $b = 1$ each f_i is a uniformly random function.*
3. *Sample a uniformly random permutation $\pi \leftarrow S_{\mathbb{F}_q \times \mathbb{F}_q}$ and output $\big(\pi \cdot \mathbf{1}_{G(f_1)}, \ldots, \pi \cdot \mathbf{1}_{G(f_m)}\big)$.*

Then D_0 and D_1 are computationally indistinguishable.

2.2 A Warm-Up Attack

For the sake of intuition, we first describe an attack on a modified version of the conjecture that is much more easily broken. Namely, we change step 3 in Restated Conjecture 1 to:

3'. Sample a uniformly random permutation $\sigma \leftarrow S_{\mathbb{F}_q}$ and output $(\sigma \cdot f_1, \ldots, \sigma \cdot f_m)$, where each f_i is interpreted as the string $(f_i(0), f_i(1), \ldots, f_i(q-1))$.

In words, the uniformly random permutation in the modified conjecture is applied to the *truth table* of the function rather than *graph* of the function. Equivalently, this modification can be viewed as only permuting the order of the columns in the graph of the function.

The Attack There is a very simple attack on the modified conjecture, distinguishing D_0 from D_1 with negligible (and one-sided) error: Given $(\sigma \cdot f_1, \ldots, \sigma \cdot f_m)$, output $b = 0$ if and only if

$$\dim\big(\operatorname{span}\{\sigma \cdot f_1, \ldots, \sigma \cdot f_m\}\big) < q. \tag{1}$$

The left-hand side is just the rank of the matrix whose i^{th} column is $\sigma \cdot f_i$, and is thus computable in polynomial time by Gaussian elimination.

The Analysis We use the standard fact that degree-λ polynomials form a $(\lambda+1)$-dimensional subspace of all functions. Thus when $b = 0$, we will always have

$$\dim\big(\operatorname{span}\{f_1, \ldots, f_m\}\big) \le \lambda + 1 < q.$$

Random functions do not have this property: for any proper subspace $V \subsetneq \mathbb{F}_q^q$, a random function $f : \mathbb{F}_q \to \mathbb{F}_q$ will lie in V with probability at most $1/q$. There are $\frac{q^q - 1}{q-1} \le q^q$ maximal (i.e., co-dimension 1) subspaces of $\mathbb{F}_q^q$, so when $f_1, \ldots, f_m$ are i.i.d. uniform, the probability that $f_1, \ldots, f_m$ are all contained in *any* subspace $V \subsetneq \mathbb{F}_q^q$ is at most $q^q \cdot q^{-m} = q^{q-m}$, which is negligible in λ by our choice of parameters. Thus when $b = 1$, we have with overwhelming probability that

$$\dim\big(\operatorname{span}\{f_1, \ldots, f_m\}\big) = q.$$

Finally, we observe that for *any* permutation σ and vectors $f_1, \ldots, f_m$, we always have

$$\dim\big(\operatorname{span}\{\sigma \cdot f_1, \ldots, \sigma \cdot f_m\}\big) = \dim\big(\operatorname{span}\{f_1, \ldots, f_m\}\big),$$

since the two sides are the ranks of matrices with identical row spaces.

2.3 The Full Attack

We now turn our attention back to Restated Conjecture 1. Our attack is most naturally presented in a more general setting:

Proposition 2.4 (Generalized Attack). *Let $V_0 \subsetneq V_1 \subseteq \mathbb{F}_q^X$ be arbitrary linear spaces parameterized by a security parameter $\lambda \in \mathbb{N}$, with q and $|X|$ polynomially bounded. Let $m = m(\lambda)$ be a sufficiently large integer ($m \ge q^2|X|\lambda$ suffices).*

For $b \in \{0,1\}$, define the distribution ensemble D_b (also parameterized by λ) as follows:

1. *Sample i.i.d. uniform $f_1, \ldots, f_m \leftarrow V_b$.*
2. *Sample a uniformly random permutation $\pi \leftarrow S_{X \times \mathbb{F}_q}$*
3. *Output $\big(\pi \cdot \mathbf{1}_{G(f_1)}, \ldots, \pi \cdot \mathbf{1}_{G(f_m)}\big)$.*

Then D_0 and D_1 are distinguishable in polynomial time. Moreover, the distinguishing error is negligible in λ and one-sided (a sample from D_0 is never mistakenly identified as coming from D_1).

Proposition 2.4 falsifies Restated Conjecture 1 with X as $\mathbb{F}_q$, V_0 as the space of degree-λ polynomials, and V_1 as all of $\mathbb{F}_q^{\mathbb{F}_q}$.

The Attack The distinguisher looks very similar to that of our warm-up attack. Given

$$\left(\pi \cdot \mathbf{1}_{G(f_1)}, \ldots, \pi \cdot \mathbf{1}_{G(f_m)}\right),$$

output $b = 0$ if and only if

$$\dim\left(\operatorname{span}\left\{\pi \cdot \mathbf{1}_{G(f_1)}, \ldots, \pi \cdot \mathbf{1}_{G(f_m)}\right\}\right) < \tau,$$

where τ is a threshold that depends on V_0, V_1. As before, the left-hand side is computable in polynomial time using Gaussian elimination.

The Analysis At a high level, we show that the distributions of $\mathbf{1}_{G(f_i)}$ in D_0 and D_1, although more complicated, retain the core properties that enabled our warm-up attack. We view each $\mathbf{1}_{G(f_i)}$ as an element of $\mathbb{F}_q^{X \times \mathbb{F}_q}$ and prove that:

- $\tilde{V}_0 \overset{\mathsf{def}}{=} \operatorname{span}\left\{\mathbf{1}_{G(f)}\right\}_{f \in V_0}$ is a proper subspace of $\tilde{V}_1 \overset{\mathsf{def}}{=} \operatorname{span}\left\{\mathbf{1}_{G(f)}\right\}_{f \in V_1}$.
- When $f \leftarrow V_1$ is uniformly random, $\mathbf{1}_{G(f)}$ is not too concentrated in any proper subspace of $\tilde{V}_1$. This implies that polynomially many samples of $\mathbf{1}_{G(f)}$ for $f \leftarrow V_1$ are very likely to span *all* of $\tilde{V}_1$, and will have higher dimension than $\tilde{V}_0$. In particular, we show in Lemma 2.8 that $\dim(\tilde{V}_1)q\lambda$ samples suffices, which is at most $q^2|X|\lambda$.

To prove these two properties, it will prove fruitful for us to first characterize functions of the form $f \mapsto \langle \mathbf{v}, \mathbf{1}_{G(f)} \rangle$ for some vector $\mathbf{v} \in \mathbb{F}^{q \cdot |X|}$.

Lemma 2.5. *Let $\mathbb{F}$ be a finite field and let $\phi \in \mathbb{F}^{\mathbb{F}^X}$ be arbitrary (that is, ϕ takes as input a function $f : X \to \mathbb{F}$ and outputs an element of $\mathbb{F}$). With f indeterminate, $\phi(f)$ can be written in the form $\langle \mathbf{v}, \mathbf{1}_{G(f)} \rangle$ if and only if it can be written in the form $\sum_{x \in X} g_x\big(f(x)\big)$ for functions $\{g_x : \mathbb{F} \to \mathbb{F}\}$.*

Proof. Let $f : X \to \mathbb{F}$ be a function. Writing $\mathbf{1}_{G(f)} = \sum_{x \in X} \mathbf{1}_{\{(x, f(x))\}}$, we see that

$$\begin{aligned} \langle \mathbf{v}, \mathbf{1}_{G(f)} \rangle &= \Big\langle \mathbf{v}, \sum_{x \in X} \mathbf{1}_{\{(x,f(x))\}} \Big\rangle \\ &= \sum_{x \in X} \langle \mathbf{v}, \mathbf{1}_{\{(x,f(x))\}} \rangle \\ &= \sum_{x \in X} v_{(x,f(x))}. \end{aligned}$$

This is equal to $\sum_x g_x\big(f(x)\big)$ if we have

$$g_x(y) = v_{(x,y)} \text{ for all } x \in X, y \in \mathbb{F}. \tag{2}$$

For any $\mathbf{v}$, one can define $\{g_x\}$ that satisfy Eq. (2), and also vice versa. □

We can now characterize the effect of the mapping $f \mapsto \mathbf{1}_{G(f)}$ on linear subspaces of $\mathbb{F}_q^X$.

Lemma 2.6. *Let $V_0 \subseteq V_1$ be linear subspaces of $\mathbb{F}^X$ for a field $\mathbb{F}$ and a set X. For $b \in \{0, 1\}$, define*

$$\tilde{V}_b \overset{\mathsf{def}}{=} \operatorname{span}\left\{\mathbf{1}_{G(f)}\right\}_{f \in V_b}$$

Then $\tilde{V}_0 = \tilde{V}_1$ if and only if $V_0 = V_1$.

Proof. Clearly if $V_0 = V_1$ then $\tilde{V}_0 = \tilde{V}_1$. Conversely, suppose $V_0 \subsetneq V_1$ and take $\mathbf{y} \in V_0^\perp \setminus V_1^\perp$. Writing $\mathbf{y} = (y_x)_{x \in X}$, let $g_x : \mathbb{F} \to \mathbb{F}$ be the function that multiplies its input by the scalar y_x. Then for any $f \in \mathbb{F}^X$ we have $\langle f, \mathbf{y} \rangle = \sum_{x \in X} g_x(f(x))$. Since $\mathbf{y} \in V_0^\perp \setminus V_1^\perp$, we have $\sum_{x \in X} g_x(f(x)) = 0$ for all $f \in V_0$, but not for all $f \in V_1$. By Lemma 2.5, we can translate this to a vector $\mathbf{v} \in \tilde{V}_0^\perp \setminus \tilde{V}_1^\perp$, which establishes that $\tilde{V}_0 \neq \tilde{V}_1$.

□

Lemma 2.7. *Let $V \subseteq \mathbb{F}_q^X$ be an arbitrary subspace, and let $\tilde{V}$ denote* $\operatorname{span}\left\{\mathbf{1}_{G(f)}\right\}_{f\in V}$. *Then for every $\tilde{U} \subsetneq \tilde{V}$,*

$$\Pr_{F\leftarrow V}\left[\mathbf{1}_{G(F)} \in \tilde{U}\right] \le \frac{q-1}{q}.$$

Proof. Let $\mathbf{u}$ be a vector in $\tilde{U}^\perp \setminus \tilde{V}^\perp$, and let $\{g_x : \mathbb{F}_q \to \mathbb{F}_q\}_{x\in X}$ be functions (given by Lemma 2.5) such that $\langle \mathbf{u}, \mathbf{1}_{G(f)}\rangle = \sum_x g_x\big(f(x)\big)$ for all $f \in \mathbb{F}_q^X$. We then have

$$\begin{aligned}\Pr_{F\leftarrow V}\left[\mathbf{1}_{G(F)} \in \tilde{U}\right] &\le \Pr_{F\leftarrow V}\left[\langle \mathbf{u}, \mathbf{1}_{G(F)}\rangle = 0\right]\\ &= \Pr_{F\leftarrow V}\Big[\sum_{x\in X} g_x\big(F(x)\big) = 0\Big]. \qquad (3)\end{aligned}$$

Note that F can be written $F = \sum_i \alpha_i \mathbf{v}_i$, where $\{\mathbf{v}_i\}$ is a basis for V and $\{\alpha_i\}$ are i.i.d. uniform on $\mathbb{F}_q$. We first show that

$$\sum_{x\in X} g_x\big(F(x)\big) \qquad (4)$$

is a non-zero polynomial in $\{\alpha_i\}$ of degree at most $q-1$. To establish the degree bound, we observe that for every x:

1. g_x has degree $q-1$, and
2. $F(x)$ is a degree-1 polynomial in $\{\alpha_i\}$.

The first claim is by Lagrange interpolation, and uses no structure of g_x other than that it is a function mapping $\mathbb{F}_q \to \mathbb{F}_q$. The second claim follows from $F(x) = \sum_i \alpha_i \mathbf{v}_i(x)$.

Finally, (4) is a *non-zero* polynomial because $\mathbf{u} \notin \tilde{V}^\perp$. Schwartz-Zippel thus bounds (3) by $\frac{q-1}{q}$ as desired. □

Lemma 2.8. *Let V be any subspace of $\mathbb{F}_q^X$ and let $d \coloneqq \dim(\tilde{V})$. Then*

$$\Pr_{f_1,\dots,f_{dq\lambda}\leftarrow V}[\operatorname{span}(\{\mathbf{1}_{G(f_i)}\}_{i\in[dq\lambda]}) = \tilde{V}] = 1 - \operatorname{negl}(\lambda).$$

Proof. For each $k \in [d]$, let $t_k \coloneqq \operatorname{span}(\{\mathbf{1}_{G(f_i)}\}_{i\in[kq\lambda]}$. For any $k < d$, by Lemma 2.7,

$$\Pr[t_{k+1} > t_k \mid t_k < d] = 1 - \left(\frac{q-1}{q}\right)^{q\lambda} = 1 - \operatorname{negl}(\lambda).$$

It follows that $t_d = d$ with probability $1 - \operatorname{negl}(\lambda)$. □

3 Context for Our Attack

Before addressing our attack, we first summarize the framework for constructing SK-DEPIR that was outlined by [BIPW17, CHR17], and we explain how Conjecture 1 fits in. Our current understanding of how these parameters affect security is quite limited.

3.1 The General Template

The main idea in [BIPW17, CHR17] is to start with a locally decodable code $C : \{0,1\}^n \to \{0,1\}^N$ and construct an SK-DEPIR scheme where the secret key is a random permutation $\pi \in S_N$ and a secret-key encryption key sk. Let λ denote the ciphertext length of this encryption scheme.

The general template is as follows:

- For a database $D \in \{0,1\}^n$, the server stores a string $\tilde{D} \in \left(\{0,1\}^\lambda\right)^N$ such that $\tilde{D}_{\pi(I)} = \mathsf{Enc}(\mathsf{sk}, C(D)_I)$ for every $I \in [N]$.
- To query the i^{th} element of D, the client:
    1. Runs the algorithm for locally decoding C at index $i \in [n]$, which produces a list of queries $I_1, \dots, I_k \in [N]$.

2. Sends $\pi(I_1), \dots, \pi(I_k)$ to the server.
3. Answers the aforementioned local decoding queries with the decryption of the server's responses.
4. Outputs whatever the local decoding algorithm outputs.

The correctness and efficiency of this construction follow as in [BIPW17, CHR17].

3.2 Reed-Muller Instantiations

It is easy to contrive locally decodable codes with which this template is insecure. Still, [BIPW17, CHR17] conjectured that appropriate Reed-Muller codes (with an appropriate local decoding procedure) lead to a secure instantiation. However, there still are many possibilities for instantiating the parameters of these codes and their corresponding local decoding algorithms. Each choice of parameters corresponds to a different candidate SK-DEPIR scheme.

Code Parameters Reed-Muller codes are parameterized by a field $\mathbb{F}$ and integers m and d (for each message length). They encode a message $X \in \{0,1\}^n$ systematically as a degree-d, m-variate polynomial $\hat{X} : \mathbb{F}^m \to \mathbb{F}$. Our current intuition is that these parameters do not directly impact security; they seem relevant primarily to efficiency, namely the tradeoff between the code's *rate* and *local decoding query complexity*. One piece of evidence in support of this is that neither our attack nor the lower bounds of [BHW19] depend on any particular choice of $\mathbb{F}$, m, and d.

Local Decoding Parameters Local decoding algorithms for Reed-Muller codes are based on the observation that for any $i \in \mathbb{F}^m$, $\hat{X}(i)$ can be recovered from $\left(\hat{X}(i')\right)_{i' \in S}$, where $S \subseteq \mathbb{F}^m$ contains any sufficiently large subset of any low-degree curve $\gamma : \mathbb{F} \to \mathbb{F}^m$ passing through i. There are three major desiderata in the selection of such a set, each of which appears to significantly affect whether the resulting SK-DEPIR scheme is secure. We recall these desiderata and possible resolutions that were proposed by [CHR17, BIPW17]:

- *What distribution should γ have?*

  There were two proposals discussed in [BIPW17, CHR17]. The first proposal was to let γ be a uniformly random degree-λ curve conditioned on $\gamma(0) = i$. The second proposal was to fix the first component of γ to the identity function, and pick the rest of the components uniformly at random conditioned on $\gamma(i_1) = (i_1, \dots, i_m)$. We refer to the first proposal as the parametric proposal, and the second proposal as the explicit proposal.

- *How many points on γ, and which ones, should be included?*

  The proposals of [BIPW17, CHR17] can be viewed as starting with a "base signal" that in the parametric case is equal to $\{\gamma(t)\}_{t \neq 0}$, and in the explicit case is equal to $\{\gamma(t)\}_{t \neq i_1} \cup \left(\{i_1\} \times \{z\}\right)$, where $z \leftarrow \mathbb{F}^{m-1}$ is chosen uniformly at random. Each point of the base signal is then included with some fixed probability α, which we refer to as the signal amplitude.

- *What "noise" points (outside the image of γ), if any, should be included?*

  The most natural noise distribution (and the one proposed in [BIPW17, CHR17]) is obtained by including non-signal points independently with some fixed probability $p \in [0,1]$. We refer to the ratio $\alpha/p \in [0, +\infty)$ as the signal-to-noise ratio, and denote it by β.

4 New Conjectures

At the time it was proposed, Conjecture 1 corresponded to a SK-DEPIR scheme whose parameter choices were most amenable to cryptanalysis. While the scheme was not known to be insecure, it was estimated to be a combination of "simple to analyze" and "least likely to be secure". In this section we propose a new toy conjecture to take the place of Conjecture 1, as well as a stronger version that would imply a full-fledged SK-DEPIR scheme. Along the way we discuss two alternative directions for SK-DEPIR.

4.1 New Toy Conjectures

There are several possible modifications to Conjecture 1 that may result in a true conjecture. Each modification that we discuss attempts to base SK-DEPIR on Reed-Muller codes, and in particular modifies the local decoding procedure described above.

Parametrism To query $i \in \mathbb{F}^m$, sample $\gamma : \mathbb{F} \to \mathbb{F}^m$ uniformly at random such that $\gamma(0) = i$ (not fixing γ_1 to be the identity function).

One drawback of this modification is that the indicator vector $\mathbf{1}_{\mathrm{Img}(\gamma)}$ is no longer λ-wise independent. Consequently the statistical query lower bound of [BHW19] does not extend to this modification. We leave it as an interesting topic for future research to determine whether there exist statistical query attacks on this variant.

Sub-sampling and Noise Instead of querying the entire image of γ, query a random subset instead of a given size. We note that using smaller subset sizes can only be more secure. On the other hand, the corresponding SK-DEPIR constructions require a minimum subset size for functionality. In addition to querying the image of γ, one can also query $\approx q$ random ("noisy") points. We note that adding noise can only improve security, as long as the distribution of noise is independent of γ.

More generally, it makes sense to combine subsampling and noise, and parameterize the resulting conjecture by the "signal amplitude" α, and the "signal-to-noise ratio" β. Setting either $\alpha < 1$ or $\beta < +\infty$ seems to eliminate the linear structure leveraged by our attack, although we lack a good way of formalizing this. We suggest that the modification most amenable to further analysis is to simply change α to $1/2$, while keeping $\beta = +\infty$.

4.2 New Conjectures for Full-Fledged SK-DEPIR

4.2.1 General Conjecture

Let λ denote a security parameter, and let $C \subseteq [q]^n$ be any linear code with a dual distance d such that:

- The block length n and the alphabet size q are both $\lambda^{O(1)}$.
- C's dual distance d is at least λ.

Furthermore, let $m = m(\lambda)$ be any polynomial in λ, and let $w = w(\lambda)$ be such that $\frac{w}{n} \leq 1 - \Omega(1)$. We conjecture that the following two distributions are computationally indistinguishable.

Structured Distribution

1. Sample a random permutation $\pi \leftarrow S_{[n]\times[q]}$.
2. Repeat m times:
   (a) Sample $(i_0, \sigma_0) \leftarrow [n] \times [q]$.
   (b) Sample $c \leftarrow C$ conditioned on $c_{i_0} = \sigma_0$. This is well-defined because C's dual distance is greater than 1, so for every i the distribution of c_i when sampling $c \leftarrow C$ is uniform on $[q]$.
   (c) Sample $\sigma_0' \leftarrow [q]$.
   (d) Define $c' \in [q]^n$ such that
   $$c'_i = \begin{cases} \sigma'_0 & \text{if } i = i_0 \\ c_i & \text{otherwise.} \end{cases}$$
   (e) Sample $i_1, \ldots, i_w \leftarrow [n]$.
   (f) Output $\big(i_0, \sigma_0, \pi(i_1, c'_{i_1}), \ldots, \pi(i_w, c'_{i_w})\big)$.

Random Distribution

1. Sample a random permutation $\pi \leftarrow S_{[n]\times[q]}$.
2. Repeat m times:
   (a) Sample $(i_0, \sigma_0) \leftarrow [n] \times [q]$.
   (b) Sample $c' \leftarrow [q]^n$.

(c) Sample $i_1, \dots, i_w \leftarrow [n]$.

(d) Output $\big(i_0, \sigma_0, \pi(i_1, c'_{i_1}), \dots, \pi(i_w, c'_{i_w})\big)$.

We emphasize that in the random distribution, the m pairs (i_0, σ_0) are independent of all other outputs, whereas in the structured distribution they are correlated through the choice of c.

4.2.2 Specific Conjecture

We propose more specifically that for every polynomial $m = m(\lambda)$, the above conjecture holds when:

- $n = q$ is a random $\log(\lambda)$-bit prime (so q is $\Theta(\lambda)$).
- The code C is the set of all (truth tables of) univariate polynomials over $\mathbb{F}_q$ with degree $q^{1/4}$.
- w is $q^{5/6}$.

A SK-DEPIR Construction This conjecture implies the existence of a SK-DEPIR scheme that, for a database of size N, requires the server to first spend $O(N^2)$ time in pre-processing, and then allows the server to answer queries with $O(N^{5/6})$ computation per query.

Specifically, in the corresponding SK-DEPIR scheme:

- There is a field $\mathbb{F}_N$ of size $\approx N$ associated with each database length N.
- The secret key is a random permutation π of $\mathbb{F}_N \times \mathbb{F}_N$ and a secret-key encryption secret key sk.
- For a database D, let it be encoded by a degree-$\sqrt{N}$ bivariate polynomial Q over $\mathbb{F}_N$, i.e. with $D \equiv Q|_{H^2}$ for some fixed subset $H \subseteq \mathbb{F}_N$ with $|H| = \sqrt{N}$. The server stores $\tilde{Q} \in \big(\{0,1\}^\lambda\big)^{\mathbb{F}_N \times \mathbb{F}_N}$ such that for all $i \in \mathbb{F}_N \times \mathbb{F}_N$, $\tilde{Q}_{\pi(i)} = \mathsf{Enc}\big(\mathsf{sk}, Q(i)\big)$.
- To query an element of D (i.e. recover $Q(x^\star, y^\star)$ for some $(x^\star, y^\star) \in H^2$), the client:
  1. Samples a degree-$N^{1/4}$ univariate polynomial p with the property that $p(x^\star) = y^\star$.
  2. Defines $\tilde{p}$ to be identical to p, but with $\tilde{p}(x^\star)$ freshly uniformly random.
  3. Queries $\tilde{Q}$ on $N^{5/6}$ points of the form $\pi^{-1}(x, \tilde{p}(x))$, where the points x are i.i.d. uniform on $\mathbb{F}_N$.
  4. Upon receiving answers (a_x), interpolates a degree-$N^{3/4}$ univariate polynomial $g : \mathbb{F}_N \to \mathbb{F}_N$ such that $g(x) = \mathsf{Dec}(\mathsf{sk}, a_x)$ for all x (not including $x^\star$).
  5. Outputs $g(x^\star)$.

The correctness and efficiency of this construction follow as in [BIPW17, CHR17].

A Security Reduction We describe how the conjecture of Section 4.2.2 implies the security of the above construction.

Suppose there is an adversary that attacks the SK-DEPIR construction. That is, the adversary is able to choose two distinct sequences of database indices $\mathbf{i}^0 = (i^0_1, \dots, i^0_\ell)$ and $\mathbf{i}^1 = (i^1_1, \dots, i^1_\ell)$ such that if the adversary sees SK-DEPIR queries for $\mathbf{i}^b$ for $b \leftarrow \{0,1\}$, then it can guess b with probability noticeably larger than $1/2$.

We construct a simulator that first gets as input a sample from one of the two distributions (structured or random) defined in Section 4.2.1, with $\lambda \approx \sqrt{N}$ and $m \geq N\ell\lambda$. This sample is a list of m entries, each of the form $(i_0, \sigma_0, \tilde{i}_1, \dots, \tilde{i}_w)$, where $(i_0, \sigma_0), \tilde{i}_1, \dots, \tilde{i}_w$ are all in $[q]^2$. It then repeatedly receives an index $i \in H^2$ from the adversary, and returns what is purportedly a SK-DEPIR query to i. Specifically, the simulator looks for an entry (that it has not used before) where $(i_0, \sigma_0) = i$. By the choice of m, there will with overwhelming probability be such an entry. It then returns the corresponding $(\tilde{i}_1, \dots, \tilde{i}_w)$.

This simulator has the property that when it initially received a sample from the structured distribution, then the simulation is faithful — the simulator's answers to the adversary's queries are distributed like those of a real client. On the other hand, when the simulator initially receives a sample from the random distribution, then the simulator's answers are independent of the indices provided by the adversary.

This gives us a distinguisher breaking the conjecture, which is a contradiction. Specifically, the distinguisher receives a sample from either the structured or random distribution, and needs to guess which. It runs the adversary, which produces database indices $\mathbf{i}^0 = (i^0_1, \dots, i^0_\ell)$ and $\mathbf{i}^1 = (i^1_1, \dots, i^1_\ell)$. It samples $b \leftarrow \{0,1\}$, and passes $\mathbf{i}^b$ to the simulator (who is initially given the same input as the distinguisher — a sample from either the structured or random distribution), and forwards the simulator's answers to the adversary. The distinguisher guesses that its input was from the structured distribution if and only if the adversary correctly guesses b.

References

[BHW19] Elette Boyle, Justin Holmgren, and Mor Weiss. Permuted puzzles and cryptographic hardness. In Dennis Hofheinz and Alon Rosen, editors, *TCC 2019, Part II*, volume 11892 of *LNCS*, pages 465–493. Springer, Heidelberg, December 2019.

[BIM00] Amos Beimel, Yuval Ishai, and Tal Malkin. Reducing the servers computation in private information retrieval: PIR with preprocessing. In Mihir Bellare, editor, *CRYPTO 2000*, volume 1880 of *LNCS*, pages 55–73. Springer, Heidelberg, August 2000.

[BIPW17] Elette Boyle, Yuval Ishai, Rafael Pass, and Mary Wootters. Can we access a database both locally and privately? In Yael Kalai and Leonid Reyzin, editors, *TCC 2017, Part II*, volume 10678 of *LNCS*, pages 662–693. Springer, Heidelberg, November 2017.

[BW21] Keller Blackwell and Mary Wootters. A note on the permuted puzzles toy conjecture, 2021.

[CGKS95] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *36th FOCS*, pages 41–50. IEEE Computer Society Press, October 1995.

[CHR17] Ran Canetti, Justin Holmgren, and Silas Richelson. Towards doubly efficient private information retrieval. In Yael Kalai and Leonid Reyzin, editors, *TCC 2017, Part II*, volume 10678 of *LNCS*, pages 694–726. Springer, Heidelberg, November 2017.

[KO00] Eyal Kushilevitz and Rafail Ostrovsky. One-way trapdoor permutations are sufficient for non-trivial single-server private information retrieval. In Bart Preneel, editor, *EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 104–121. Springer, Heidelberg, May 2000.